A Unified Generative Framework for Various NER Subtasks

Abstract

Named Entity Recognition (NER) is the task of identifying spans that represent entities in sentences. Whether the entity spans are nested or discontinuous, the NER task can be categorized into the flat NER, nested NER, and discontinuous NER subtasks. These subtasks have been mainly solved by the token-level sequence labelling or span-level classification. However, these solutions can hardly tackle the three kinds of NER subtasks concurrently. To that end, we propose to formulate the NER subtasks as an entity span sequence generation task, which can be solved by a unified sequence-to-sequence (Seq2Seq) framework. Based on our unified framework, we can leverage the pre-trained Seq2Seq model to solve all three kinds of NER subtasks without the special design of the tagging schema or ways to enumerate spans. We exploit three types of entity representations to linearize entities into a sequence. Our proposed framework is easy-to-implement and achieves state-of-the-art (SoTA) or near SoTA performance on eight English NER datasets, including two flat NER datasets, three nested NER datasets, and three discontinuous NER datasets.